Datasets Created in VISCERAL

Krenn, M ; Grünberg, K ; Jimenez-del-Toro, O ; Jakab, András ; Salas Fernandez, T ; Winterstein, M ; Weber, M A ; Langs, G

DOI: https://doi.org/10.1007/978-3-319-49644-3

Chapter 5
Datasets Created in VISCERAL

Markus Krenn, Katharina Grünberg, Oscar Jimenez-del-Toro, András Jakab, Tomàs Salas Fernandez, Marianne Winterstein, Marc-André Weber and Georg Langs

Abstract In the VISCERAL project, several *Gold Corpus* datasets containing medical imaging data and corresponding manual expert annotations have been created. These datasets were used for training and evaluation of participant algorithms in the VISCERAL Benchmarks. In addition to Gold Corpus datasets, the architecture of VISCERAL enables the creation of *Silver Corpus* annotations of far larger datasets, which are generated by the collective ensemble of submitted algorithms. In this chapter, three Gold Corpus datasets created for the VISCERAL Anatomy, Detection and Retrieval Benchmarks are described. Additionally, we present two datasets that have been created as a result of the anatomy and retrieval challenge.

Source code is available at:
https://github.com/Visceral-Project/silverCorpusFramework

M. Krenn (✉) · G. Langs · A. Jakab
Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria
e-mail: markus.krenn@meduniwien.ac.at; max@contentflow.com

G. Langs
e-mail: georg.langs@meduniwien.ac.at

K. Grünberg · M. Winterstein · M.-A. Weber
University of Heidelberg, Heidelberg, Germany
e-mail: katharina.gruenberg@med.uni-heidelberg.de

T. Salas Fernandez
Agencia D'Informació, Avaluació I Qualitat En Salut, Barcelona, Spain

O. Jimenez-del-Toro
University of Applied Sciences Western Switzerland, Sierre, Switzerland
e-mail: oscar.jimenez@hevs.ch

A. Hanbury et al. (eds.), *Cloud-Based Benchmarking of Medical Image Analysis*, DOI 10.1007/978-3-319-49644-3_5

5.1 Introduction

One of the main objectives of the VISCERAL project has been to provide substantial Gold Corpus datasets to the medical image analysis research community containing medical imaging data complemented with manual annotations performed by experienced radiologists. For each benchmark organized within the project, a Gold Corpus dataset was created in order to train and evaluate the participants' algorithms.

In addition to the Gold Corpus of expert-annotated imaging data, the architecture of the VISCERAL Benchmarks offers the possibility to generate far larger Silver Corpus data that are annotated by the collective ensemble of algorithms submitted by Benchmark participants. Even though these Silver Corpus annotations are expected to be less accurate than Gold Corpus annotations, we encourage the idea of their creation since they can be generated automatically and therefore created on larger scales than is feasible to achieve with expert annotations. Furthermore, experiments showed that the pooling of algorithm results did provide enhanced annotations over individual algorithms [1].

The following sections describe Gold Corpus and Silver Corpus datasets created as part of VISCERAL.

5.2 Anatomy Gold Corpus

The Anatomy Gold Corpus was created to provide substantial training and test data for the Anatomy Benchmarks 1–3, in which participants have been challenged with the tasks of labelling anatomical structures (segmentation) on the one hand and detecting landmarks (localization) in medical imaging data on the other hand.

The dataset contains 120 3D medical images (volumes) acquired during daily clinical routine and cover four different imaging modalities. Table 5.1 lists and describes the modalities, their fields of view and voxel dimensions.

Each volume carries two types of anatomical reference annotations performed by experienced radiologists that serve as gold standard references:

Table 5.1 Imaging modalities covered by the VISCERAL Anatomy Gold Corpus

Identifier	Modality	Field of view	Voxel dimensions (in mm)
CT-Wb	CT	whole body	$0.8 - 0.9 \times 0.8 - 0.9 \times 1.5$
CTce-ThAb	contrast-enhanced CT	thorax and abdomen	$0.6 - 0.7 \times 0.6 - 0.7 \times 1.2 - 1.5$
MRT1-Wb	MRI - T1 weighted	whole body	$1.1 - 1.3 \times 1.1 - 1.3 \times 6 - 7$
MRT1cefs-Ab	contrast-enhanced fat-saturated MRI - T1 weighted	abdomen	$1.2 - 1.3 \times 1.2 - 1.3 \times 3$

Table 5.2 Overview of the VISCERAL Anatomy Gold Corpus

Category	# Volumes	# Structures	# Landmarks
CT-Wb	30	573	1574
CTce-ThAb	30	583	1244
MRT1-Wb	30	442	1447
MRT1-ThAb	30	322	595
$\sum$	**120**	**1920**	**4860**

1. **Segmentation labels**: A labelling of up to 20 anatomical structures such as kidneys, lungs, liver, urinary bladder, pancreas, adrenal glands, thyroid glands, aorta and some muscles.
2. **Landmark labels**: Up to 53 anatomical landmarks including the lateral end of the clavicula, crista iliaca, symphysis, trochanter major/minor, tip of aortic arch, trachea/aortic bifurcation, crista iliaca and the vertebrae.

An anatomical structure annotation is given in the form of a 3D image, where the value 0 in a voxel indicates absence (background) and a value > 0 indicates presence (foreground) of a specific structure. All annotated landmarks of an image are given as a list where an entry holds the landmark name and its x-, y- and z- coordinates.

In VISCERAL, the Neuroimaging Informatics Technology Initiative (NIfTI)[1] file format is used to store medical imaging data. In contrast to the slice-based Digital Imaging and Communications in Medicine (DICOM)[2] standard, the full volume is stored as a single self-contained file. This facilitates file management considerably, since transfer and storage of thousands of large files instead of millions of small files are typically more efficient.

Table 5.2 lists the number of volumes, annotated structures and landmarks that build the VISCERAL Anatomy Gold Corpus. Overall, 30 volumes of each modality have been annotated, resulting in a dataset that consists of 120 volumes with 1920 corresponding structures and 4860 landmark annotations. Detailed breakdowns of annotations per structure and landmark in each modality are given in Tables 5.3 and 5.4, where Table 5.3 provides a breakdown of manually annotated anatomical structures per modality in volumes of the Anatomy Gold Corpus dataset, and Table 5.4 lists landmark annotations that have been annotated by radiology experts and are available in volumes of the Gold Corpus. Missing annotations are due to poor visibility of the structures in certain image modalities or due to such structures being outside of the field of view. Figure 5.1 illustrates Gold Corpus annotations in one volume of each modality.

[1] Neuroimaging Informatics Technology Initiative: http://nifti.nimh.nih.gov/.

[2] Digital Imaging and Communications in Medicine: http://dicom.nema.org/.

Table 5.3 Manual annotations of anatomical structures performed by experienced radiologists available in the Anatomy Gold Corpus

Structure	CT-Wb	Ctce-ThAb	MRT1-Wb	MRT1cefs-Ab	$\sum$
Adrenal gland (L)	24	28	17	16	**85**
Adrenal gland (R)	21	28	14	8	**71**
Aorta	30	30	30	10	**100**
First lumbar vertebra	30	30	29	22	**111**
Gallbladder	25	29	9	14	**77**
Kidney (L)	29	30	30	28	**117**
Kidney (R)	30	30	30	28	**118**
Liver	30	30	28	30	**118**
Lung (L)	30	30	30	7	**97**
Lung (R)	30	30	30	7	**97**
M. b. rectus abdominis (L)	30	30	4	7	**71**
M. b. rectus abdominis (R)	30	30	4	6	**70**
Pancreas	30	28	9	21	**88**
Psoas major (L)	30	30	30	29	**119**
Psoas major (R)	30	30	30	30	**120**
Spleen	30	30	30	29	**119**
Sternum	30	30	7		**67**
Thyroid gland	25	20	21		**66**
Trachea	30	30	30		**90**
Urinary bladder	29	30	30	30	**119**
$\sum$	**573**	**583**	**442**	**322**	**1920**

Table 5.4 Annotated landmarks per modality available in volumes of the Anatomy Gold Corpus

Landmark	CT-Wb	Ctce-ThAb	MRT1-Wb	MRT1cefs-Ab	$\sum$
Aorta bifurcation	30	30	29	30	**119**
Aortic arch	30	30	29	2	**91**
Aortic valve	29	30	24		**83**
Bronchus (L)	30	28	25		**83**
Bronchus (R)	30	28	27		**85**
C2	28		29		**57**
C3	29		29		**58**
C4	29		29		**58**
C5	29		29		**58**
C6	29	6	29		**64**
C7	29	22	29		**80**
Clavicle (L)	30	13	30	30	**103**
Clavicle (R)	30	13	30	30	**103**

(continued)

Table 5.4 (continued)

Landmark	CT-Wb	Ctce-ThAb	MRT1-Wb	MRT1cefs-Ab	∑
Coronaria	23	22	1		**46**
Crista iliaca (L)	30	30	30		**90**
Crista iliaca (R)	30	30	30		**90**
Eye (L)	30		5		**35**
Eye (R)	30		5		**35**
Ischiadicum (L)	30	30	29	24	**113**
Ischiadicum (R)	30	30	29	24	**113**
L1	30	30	30	30	**120**
L2	30	30	30	30	**120**
L3	30	30	30	30	**120**
L4	30	31	30	30	**121**
L5	30	30	30	30	**120**
Renalpelvis (L)	29	30	29	27	**115**
Renalpelvis (R)	30	30	30	27	**117**
Sternoclavicular (L)	30	30	27		**87**
Sternoclavicular (R)	30	30	27		**87**
Symphysis	30	30	29	30	**119**
Th1	30	30	30		**90**
Th2	30	30	31		**91**
Th3	30	30	28		**88**
Th4	30	30	29		**89**
Th5	30	30	29		**89**
Th6	30	30	29	1	**90**
Th7	30	30	30	3	**93**
Th8	30	30	30	5	**95**
Th9	30	30	30	9	**99**
Th10	30	30	30	19	**109**
Th11	30	30	30	25	**115**
Th12	30	30	29	28	**117**
Trachea bifurcation	30	29	29		**88**
Trochanter major (L)	30	30	30	22	**112**
Trochanter major (R)	30	30	30	24	**114**
Trochanter minor (L)	30	29	30	20	**109**
Trochanter minor (R)	30	29	30	20	**109**
Tuberculum (L)	30	17	30		**77**
Tuberculum (R)	30	17	30		**77**
Vci bifurcation	30	30	27		**87**
Ventricle (L)	30		29		**59**
Ventricle (R)	30		29	30	**89**
Xyphoideus	30	30	9	15	**84**
∑	**1574**	**1244**	**1447**	**595**	**4860**

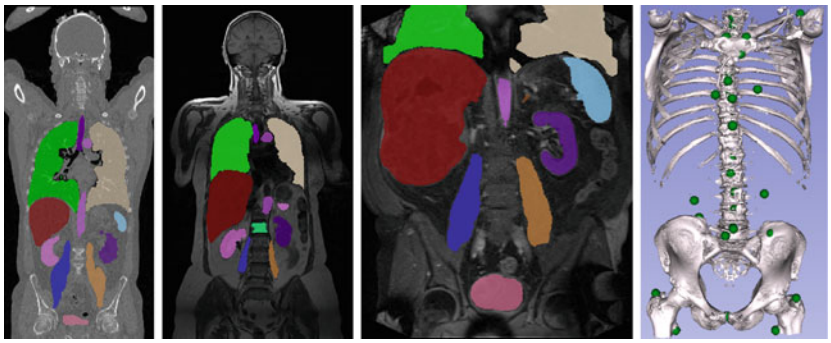

Fig. 5.1 Anatomical structure annotations in CT, MRT1 and a MRT1cefs volume *(a-c)* and visualization of annotated landmarks in a CTce volume *(d)*

5.3 Anatomy Silver Corpus

The Anatomy Silver Corpus was created based on the data and results available from the segmentation tasks of Anatomy 2 and 3 Benchmarks [2, 3]. Here, segmentations of all organs addressed within the benchmark were created by fusing multiple segmentation estimates originating from (1) the submitted algorithms and (2) Gold Corpus annotations transformed by medical image registration methods. The process to derive a Silver Corpus annotation of a specific structure in a novel volume is described and discussed in detail in [1] and can be summarized as follows:

1. Compute algorithmic segmentation estimates by applying all submitted algorithms to the target image.
2. Transfer manual annotations of Gold Corpus volumes to the target image by a preregistration selection, image registration and label propagation approach.
3. Build consensus of all segmentation estimates (algorithmic and atlas based) using the SIMPLE [4] segmentation approach.

This procedure has been applied to 264 additional volumes of the modalities covered by the Gold Corpus, resulting in up to 20 automatically generated Silver

Table 5.5 Overview of the VISCERAL Anatomy Silver Corpus dataset

Category	# Volumes	# Structures	# Landmarks
CT-Wb	62	1122	3169
CTce-ThAb	65	1227	2600
MRT1-Wb	66	1095	3136
MRT1-ThAb	71	879	1342
$\sum$	**264**	**4323**	**10247**

Table 5.6 Segmentation accuracy (μ, σ) of the silver corpus fusion process evaluated on 10 volumes of the Gold Corpus per modality and number of Silver Corpus annotations (#) computed on additional volumes that are available as a resource for the research community

Anatomical structure	**CT**			**CTce**			**MRT1**			**MRT1cefs**		
	#	μ	σ	#	μ	σ	#	μ	σ	#	μ	σ
Adrenal gland (L)	54	0.36	0.19	53	0.35	0.17	41	0.17	0.22	49	0.21	0.12
Adrenal gland (R)	54	0.32	0.2	56	0.35	0.14	50	0.38	0.14	60	0.23	0.11
Aorta	58	0.79	0.04	63	0.82	0.05	65	0.73	0.07	71	0.68	0.02
First lumbar vertebra	57	0.67	0.36	63	0.68	0.34	58	0.46	0.25	71	0.23	0.12
Gallbladder	40	0.24	0.19	49	0.54	0.15	46	0.05	0.05	61	0.13	0.2
Kidney (L)	58	0.9	0.03	63	0.93	0.02	64	0.84	0.06	71	0.85	0.2
Kidney (R)	57	0.87	0.12	63	0.94	0.01	65	0.81	0.11	71	0.86	0.18
Liver	59	0.93	0.01	63	0.94	0.01	66	0.83	0.07	71	0.9	0.03
Lung (L)	61	0.97	0.01	63	0.97	0.01	66	0.91	0.03	-	-	-
Lung (R)	60	0.98	0.01	64	0.97	0.01	66	0.92	0.02	-	-	-
M. b. rectus abdominis (L)	55	0.64	0.14	64	0.63	0.17	-	-	-	-	-	-
M. b. rectus abdominis (R)	56	0.6	0.21	63	0.69	0.16	-	-	-	-	-	-
Pancreas	57	0.43	0.19	60	0.47	0.18	63	0.21	0.21	71	0.46	0.13
Psoas major (L)	56	0.84	0.02	63	0.85	0.05	65	0.82	0.06	71	0.8	0.05
Psoas major (R)	58	0.84	0.02	63	0.86	0.02	65	0.79	0.06	71	0.73	0.12
Spleen	55	0.89	0.06	63	0.89	0.07	65	0.74	0.11	71	0.79	0.18
Sternum	55	0.8	0.04	63	0.83	0.07	64	0.6	0	-	-	-
Thyroid gland	57	0.57	0.1	62	0.52	0.13	64	0.25	0.15	-	-	-
Trachea	57	0.93	0.02	62	0.93	0.02	63	0.78	0.1	-	-	-
Urinary bladder	58	0.76	0.15	64	0.86	0.06	59	0.66	0.28	70	0.45	0.25
$\sum$	**1122**			**1227**			**1095**			**879**		

Corpus segmentations per volume. In addition to the segmentation of organs, each volume is complemented with manually performed landmark annotations similar to those of the Gold Corpus. This results in the VISCERAL Anatomy Silver Corpus that contains over 4.000 automatically generated segmentations of anatomical structures and more than 10.000 annotated landmarks.

Table 5.5 outlines the number of volumes, structure segmentations and landmark annotations in each modality available in the Silver Corpus. Detailed breakdowns of segmentations per structure and landmarks for each modality are given in Tables 5.6 and 5.7. Table 5.6 lists the number of computed segmentations (#) per structure and modality together with average segmentation performances (μ) and corresponding standard deviations (σ) of Silver Corpus segmentations computed and compared to Gold Corpus annotations of 40 volumes. These results serve as structure- and modality-specific segmentation performance estimates of generated Silver Corpus annotations. Table 5.7 lists annotated landmarks per modality of the VISCERAL Anatomy Silver Corpus.

For reference, Fig. 5.2 shows average Dice coefficients [5] obtained by comparing Silver Corpus segmentations computed in 10 volumes per modality of the Gold

Table 5.7 Annotated landmarks per modality of the Anatomy Silver Corpus

Landmark	CT-Wb	Ctce-ThAb	MRT1-Wb	MRT1cefs-Ab	$\sum$
Aorta bifurcation	62	63	64	70	**259**
Aortic arch	51	57	54		**162**
Aortic valve	48	57	34		**139**
Bronchus (L)	62	63	51		**176**
Bronchus (R)	62	63	55		**180**
C2	61	3	65		**129**
C3	62	3	65		**130**
C4	62	3	65		**130**
C5	62	4	65		**131**
C6	62	13	65		**140**
C7	62	52	65		**179**
Clavicle (L)	62	20	65		**147**
Clavicle (R)	62	22	64		**148**
Coronaria	12	36	1		**49**
Crista iliaca (L)	62	61	63	70	**256**
Crista iliaca (R)	62	61	64	70	**257**
Eye (L)	63		23		**86**
Eye (R)	61		23		**84**
Ischiadicum (L)	62	62	65	54	**243**
Ischiadicum (R)	62	62	63	54	**241**

(continued)

Table 5.7 (continued)

Landmark	CT-Wb	Ctce-ThAb	MRT1-Wb	MRT1cefs-Ab	$\sum$
L1	62	63	65	68	**258**
L2	62	63	65	70	**260**
L3	62	63	64	71	**260**
L4	61	63	65	71	**260**
L5	60	63	63	71	**257**
Renalpelvis (L)	61	62	64	69	**256**
Renalpelvis (R)	61	62	64	66	**253**
Sternoclavicular (L)	62	63	59		**184**
Sternoclavicular (R)	62	63	59		**184**
Symphysis	62	64	64	68	**258**
Th1	62	63	65		**190**
Th2	62	63	65		**190**
Th3	62	64	65		**191**
Th4	62	63	65		**190**
Th5	61	63	65		**189**
Th6	62	63	65		**190**
Th7	62	63	65		**190**
Th8	62	63	65	8	**198**
Th9	62	63	65	15	**205**
Th10	62	63	65	33	**223**
Th11	62	63	65	50	**240**
Th12	62	63	65	60	**250**
Trachea bifurcation	62	62	64		**188**
Trochanter major (L)	62	64	65	60	**251**
Trochanter major (R)	62	64	65	59	**250**
Trochanter minor (L)	61	63	64	52	**240**
Trochanter minor (R)	61	62	64	52	**239**
Tuberculum (L)	61	31	61		**153**
Tuberculum (R)	62	38	63		**163**
Vci bifurcation	60	63	64	65	**252**
Ventricle (L)	48		62		**110**
Ventricle (R)	48		63		**111**
Xyphoideus	60	62	10	16	**148**
$\sum$	**3169**	**2600**	**3136**	**1342**	**10247**

Corpus to the corresponding manual ground truth annotation. These results can be interpreted as structure and modality-specific segmentation performance estimates of generated Silver Corpus annotations. Average segmentation accuracy (μ) and

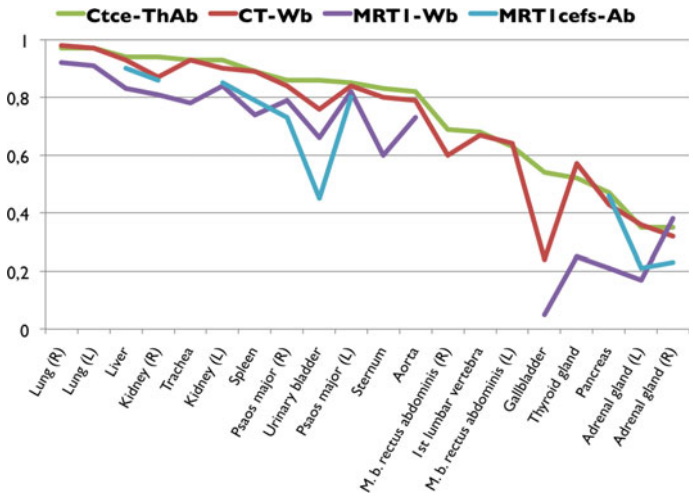

Fig. 5.2 Accuracy (DICE) of Silver Corpus segmentations evaluated on 10 volumes of the Anatomy Gold Corpus

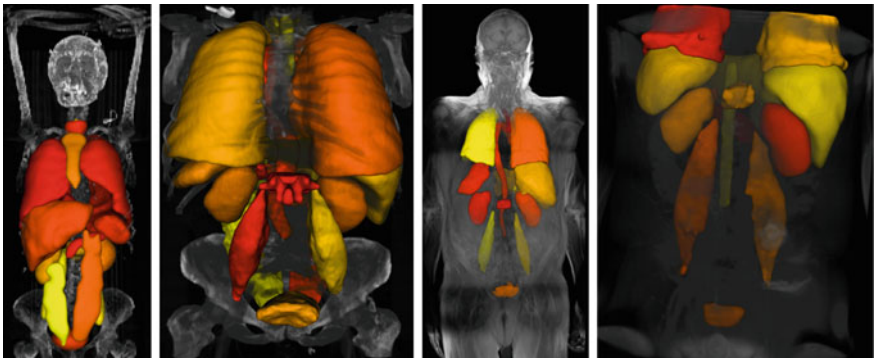

Fig. 5.3 Illustrations of generated Silver Corpus annotations in one volume of each modality. Figures taken from [1]

corresponding standard deviations (σ) are provided in Table 5.6. Figure 5.3 illustrates computed Silver Corpus segmentations in one volume of each modality. The software for creating the Silver Corpus is available.[3]

[3] https://github.com/Visceral-Project/silverCorpusFramework.

5.4 Detection Gold Corpus

In the VISCERAL Detection Benchmark, participants have been challenged to develop algorithms that automatically detect and identify lesions in medical imaging data. The Gold Corpus created for test and training purposes in this context thus consists of a set of medical images in which lesions have been manually annotated by the experienced radiologists.

The dataset includes volumes of two modalities (CT-Wb & MRT2-Wb) in which all lesions of five predefined target structures (bones, brain, liver, lung and lymph nodes) have been annotated. A lesion is identified by one point that indicates the centre of a lesion and two additional points on the perimeter to give an estimate of the diameter. Since lesions are not spherical, this is an estimate, but in the context of the Detection Benchmark still is clinically relevant. All lesion annotations of a volume are given in an *fcsv* file containing a list of annotated points and their x-, y- and z-coordinates labelled according to the following naming convention:

structure_counter_identifier, where

- *structure* indicates in which anatomical structure the lesion is located (bones BO, brain BR, liver LI, lungs LU and lymph nodes LN),
- *counter* depicts the index of a lesion within an anatomical structure and
- *identifier* defines if the annotated point represents the centre (C) or diameter estimate (D1, D2) of a specific lesion.

Figure 5.4 gives an example of a lesion annotation file and shows the three points (C, D1 and D2) that represent a bone lesion annotation in a MRT2 image. In total, 1609 lesions have been annotated in 100 volumes. Table 5.8 gives an overview of volumes and lesions annotated per modality and target structure. Example lesion annotations in all target structures of both modalities are shown in Fig. 5.5.

```
# VISCERAL - Detection Benchmark
# Lesion annotation file
#
# Each row contains one point of
# a lesion annotation
# Each lesion has three annotated points:
#  The centre of the lesion (C) and
#  2 points to estimate the diameter (D1,D2)
#
# Each annotation is named according
# to the following convention
#  <targetStructure>_<counter>_<pointIdentifier>
#
# columns = lesionIdentifier,x,y,z
#
#
BO_1_C,118.779,6.168,-657.02
BO_1_D1,118.401,6.168,-651.99
BO_1_D2,117.834,6.168,-663.44
BO_2_C,128.098,18.1587,-698.30
BO_2_D1,128.253,18.1587,-695.20
BO_2_D2,127.633,18.1587,-700.94
```

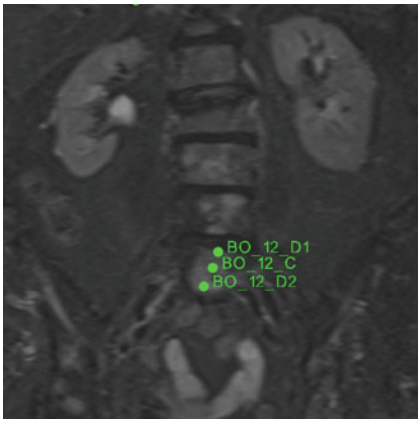

Fig. 5.4 Example of a lesion annotation file and illustration of an annotated bone lesion

Table 5.8 Number of volumes and lesions annotated in the VISCERAL detection Gold Corpus

Modality	# Volumes	# Annotated lesions					
		Bone	Lungs	Liver	Lymph nodes	Brain	∑
CT - WB	50	911	24	27	48	2	1012
MRT2 - WB	50	541	5	44	1	6	597
∑	**100**	**1452**	**29**	**71**	**49**	**8**	**1609**

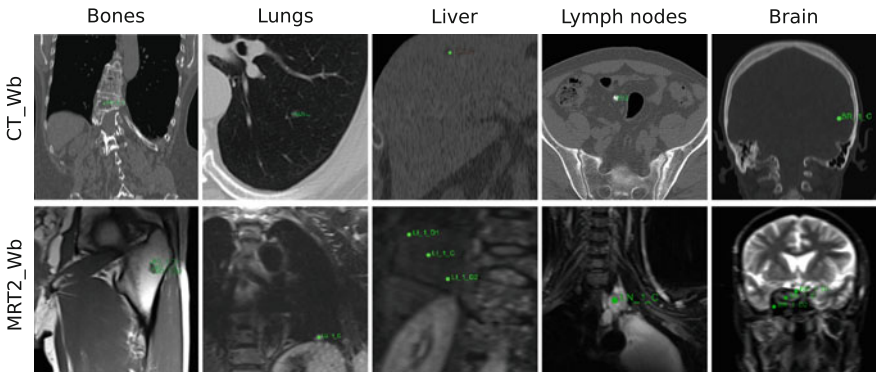

Fig. 5.5 Exemplary lesion annotations in all target structures of both modalities annotated

5.5 Retrieval Gold Corpus

Participants of the VISCERAL Retrieval Benchmark were challenged to find clinically relevant or similar cases to a given query case in a large multimodal dataset. For this purpose, a Gold Corpus has been created that contains:

1. Medical images from multiple modalities, covering different parts of the human body.
2. Anatomy–Pathology (AP) terms exported from corresponding radiology reports that describe which pathological findings occur in which anatomical regions of an image.

Annotations of findings in an image are given in the form of AP term files that list terms describing pathologies that occur in the radiology report of an image together with its anatomy. Both entities are described textually and with their corresponding RadLex ID[4] (RID). RadLex is a unified terminology of radiology terms that can be used for standardized indexing and retrieval of radiology information resources. AP term files furthermore indicate whether a pathology has been explicitly negated in the report. Figure 5.6 shows an example of an AP term file. This file indicates for instance

[4] http://rsna.org/RadLex.aspx.

```
VolumeID_Modality_Bodyregion.fcsv
i.e. 123456_MRT1_Ab.csv
```

```
# The first row depicts the header of an a-p term file.
Anatomy RID,Anatomy,Pathology RID,Pathology,Negated
#
# Each row entry in this file contains an occurring pathology and its anatomy and
states if the pathology is negated.
RID199,Ductus choledochus,RID4865,Ödem,1
RID58,Leber,RID3874,Raumforderung,0
```

Fig. 5.6 Example of an AP term file

Table 5.9 Number of volumes and available AP term files of the VISCERAL retrieval Gold Corpus

Modality	Field of view	# Volumes	# AP term files
CT	Abdomen	336	213
	Thorax & Abdomen	86	86
	Thorax	971	699
	Unknown	211	211
	Whole body	410	410
MRT1	Abdomen	167	114
	Unknown	24	24
MRT2	Abdomen	68	18
	Unknown	38	38
$\sum$		**2311**	**1813**

that volume *123456_MRT1_Ab* does not contain the pathological finding *Oedem* in *Ductus choledochus* but contains *Raumforderung* in the anatomical structure *Leber*.

The dataset consists of 2311 volumes originated from three different modalities (CT, MRT1, and MRT2) which have been acquired during clinical routine. For 1813 cases of the dataset, AP term files are available and thus part of the retrieval Gold Corpus. Table 5.9 gives a detailed overview of the number of volumes per modality and field of view and lists available AP term files that form the VISCERAL Retrieval Gold Corpus.

5.6 Retrieval Silver Corpus

Participants of the VISCERAL Retrieval Benchmark have been challenged to find clinically relevant cases in the Retrieval Gold Corpus for given queries. For this purpose, ten query cases (illustrated in Fig. 5.7) have been created, where each query in this scenario has been defined by:

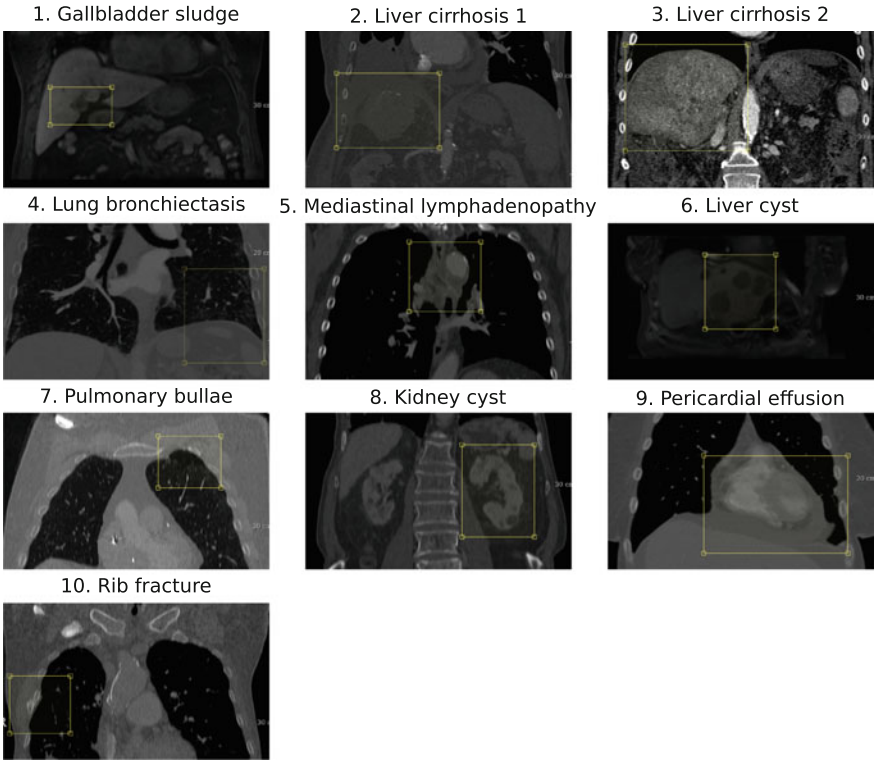

Fig. 5.7 Illustration of the query cases of the Detection Benchmark

- The AP term that defines the topic of a query, i.e. liver – cyst,
- The 3D medical image data (CT, MRT1 and MRT2),
- A 3D bounding box of the region that contains radiological signs of the pathology,
- A binary mask of the organ affected and
- The AP term list extracted from the volumes report.

During evaluation, medical experts performed relevance judgements of the top-ranked cases submitted to each query to judge the quality of retrieval of each participant's approach. This process results in a set of clinically relevant and irrelevant cases from the Gold Corpus for each given query, which builds the VISCERAL Retrieval Silver Corpus.

In total, 6240 relevance judgements have been performed in this context from which 2462 cases are clinically relevant and 3778 are not relevant to one of the given queries. Table 5.10 shows the corresponding numbers of relevant and not relevant cases of the Gold Corpus for each query.

Table 5.10 Retrieval Silver Corpus. Number of clinically relevant and not relevant cases of the Gold Corpus for each query

Query	Relevant	Not relevant	$\sum$
Gallbladder sludge	118	194	312
Liver cirrhosis 1	428	395	823
Liver cirrhosis 2	428	395	823
Lung bronchiectasis	161	453	614
Mediastinal lymphadenopathy	248	342	590
Liver cyst	339	264	603
Pulmonary bullae	333	258	591
Kidney cyst	336	263	599
Pericardial effusion	24	696	720
Rib fracture	47	518	565
$\sum$	**2462**	**3778**	**6240**

5.7 Summary

During the VISCERAL project, we have generated datasets of medical imaging data together with annotations. The purpose of the VISCERAL Gold Corpora is to serve as training set for algorithm development and for evaluation of algorithms. The VISCERAL Silver Corpora use the results of algorithms to create algorithmic annotations on far larger datasets.

Three so-called *Gold Corpus* datasets have been created containing medical imaging data and corresponding gold standard annotations:

1. The **VISCERAL Anatomy Gold Corpus** consists of 120 medical images of four modalities and carries (1) 1920 voxel-wise annotations of up to 20 anatomical structures per volume and (2) 4860 annotated landmarks of up to 53 predefined points of interest per volume.
2. The **VISCERAL Detection Gold Corpus** contains 100 medical images of two modalities and provides annotations of 1609 lesions in five anatomies (bones, lungs, liver lymph nodes and brain).
3. The **VISCERAL Retrieval Gold Corpus** includes 2311 medical images of three modalities, where for 1813 cases the corresponding radiology report-extracted AP terms are available that describe occurring pathological findings and their anatomy.

Furthermore, two *Silver Corpus* datasets have been generated based on the data and results available from Anatomy and Retrieval Benchmarks:

1. The **VISCERAL Anatomy Silver Corpus** provides automatically generated silver standard segmentations of up to 20 anatomical structures in 264 volumes (4323 in total) and additionally manual landmark annotations in each of these volumes ($>$10.000 annotations).

2. The **VISCERAL Retrieval Silver Corpus** provides a list of relevant and irrelevant cases of the Retrieval Gold Corpus to each query case of the Retrieval Benchmark.

Acknowledgements The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement 318068 (VISCERAL).

References

1. Krenn M, Dorfer M, Jiménez del Toro OA, Müller H, Menze B, Weber MA, Hanbury A, Langs G (2016) Creating a large-scale silver corpus from multiple algorithmic segmentations. In: Menze B, Langs G, Montillo A, Kelm M, Müller H, Zhang S, Cai W, Metaxas D (eds) MCV 2015. LNCS, vol 9601. Springer, Cham, pp 103–115. doi:10.1007/978-3-319-42016-5_10
2. Göksel O, Jiménez-del Toro OA, Foncubierta-Rodríguez A, Müller H (2015) Overview of the visceral challenge at ISBI 2015. In: Proceedings of the VISCERAL challenge at ISBI, New York
3. Jiménez del Toro O, Goksel O, Menze B, Müller H, Langs G, Weber M, Eggel I, Gruenberg K, Holzer M, Jakab A, Kotsios-Kontokotsios G, Krenn M, Salas Fernandez T, Schaer R, Taha AA, Winterstein M, Hanbury A (2014) Visceral—visual concept extraction challenge in radiology: ISBI 2014 challenge organization. In: Göksel O (ed) Proceedings of the VISCERAL challenge at ISBI, CEUR workshop proceedings, pp 6–15
4. Langerak TR, van der Heide UA, Kotte ANTJ, Viergever MA, van Vulpen M, Pluim JPW (2010) Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple). IEEE Trans Med Imaging 29(12):2000–2008
5. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26(3):297–302